On the Robustness of NK-Kauffman Networks Against Changes in their Connections and Boolean Functions

Federico Zertuche

Instituto de Matemáticas, Unidad Cuernavaca
Universidad Nacional Autónoma de México
A.P. 273-3, 62251 Cuernavaca, Morelos., México.
zertuche@matcuer.unam.mx

Abstract

NK-Kauffman networks $\mathcal{L}_K^N$ are a subset of the Boolean functions on N Boolean variables to themselves, $\Lambda_N = \left\{\xi : \mathbb{Z}_2^N \to \mathbb{Z}_2^N\right\}$. To each NK-Kauffman network it is possible to assign a unique Boolean function on N variables through the function $\Psi : \mathcal{L}_K^N \to \Lambda_N$. The probability $\mathcal{P}_K$ that $\Psi(f) = \Psi(f')$, when f' is obtained through f by a change of one of its K-Boolean functions ($b_K : \mathbb{Z}_2^K \to \mathbb{Z}_2$), and/or connections; is calculated. The leading term of the asymptotic expansion of $\mathcal{P}_K$, for $N \gg 1$, turns out to depend on: the probability to extract the *tautology* and *contradiction* Boolean functions, and in the average value of the distribution of probability of the Boolean functions; the other terms decay as $\mathcal{O}(1/N)$. In order to accomplish this, a classification of the Boolean functions in terms of what I have called their *irreducible degree of connectivity* is established. The mathematical findings are discussed in the biological context where, Ψ is used to model the genotype-phenotype map.

Short title: *Robustness Against Mutations.*

Keywords: *Cellular automata, irreducible connectivity, binary functions, functional graphs, redundant genetic material, genetic robustness.*

PACS numbers: 05.65.+b, 87.10.+e, 87.14.Gg, 89.75.Fb

1. Introduction

NK-Kauffman networks are useful models for the study the genotype-phenotype map Ψ; which is their main application in this work [1,2]. An NK-Kauffman network consists of N Boolean variables $S_i(t) \in \mathbb{Z}_2$ $(i = 1, \ldots, N)$, that evolve deterministically in discretized time $t = 0, 1, 2, \ldots$ according to Boolean functions on K $(0 \leq K \leq N)$ of these variables at the previous time $t - 1$. For every site i, a K-Boolean function $f_i : \mathbb{Z}_2^K \to \mathbb{Z}_2$ is chosen randomly and independently with a bias probability p $(0 < p < 1)$, that $f_i = 1$ for each of its possible 2^K arguments. Also, for every site i, K inputs (the connections) are randomly selected from a uniform distribution, among the N Boolean variables of the network, without repetition. Once the K inputs and the functions f_i have been selected, a Boolean deterministic dynamical system; known as a NK-Kauffman network has been constructed. The network evolves deterministically, and synchronously in time, according to the rules

$$S_i(t+1) = f_i\left(S_{i_1}(t), S_{i_2}(t), \ldots, S_{i_K}(t)\right), \quad i = 1, \ldots, N, \tag{1}$$

where $i_m \neq i_n$, for all $m, n = 1, 2, \ldots, K$, with $m \neq n$; because all the inputs are different. An NK-Kauffman network is then a map of the form

$$f : \mathbb{Z}_2^N \longrightarrow \mathbb{Z}_2^N.$$

Let us denote by $\mathcal{L}_K^N$ the set of NK-Kauffman networks that might be built up, by this procedure, for given N and K. They constitute a subset of the set of all possible Boolean functions on N-Boolean variables to themselves

$$\Lambda_N = \left\{\xi : \mathbb{Z}_2^N \to \mathbb{Z}_2^N\right\}.$$

In Ref. 1, a study of the injective properties of the map

$$\Psi : \mathcal{L}_K^N \to \Lambda_N \tag{2}$$

was pursuit for the case $p = 1/2$; where the Boolean functions are extracted from a uniform distribution. Using the fact that $\Lambda_N \cong \mathcal{G}_{2^N}$, where $\mathcal{G}_{2^N}$ is the set of functional graphs on 2^N points [3]; the average number $\vartheta(N, K)$ of elements in $\mathcal{L}_K^N$ that Ψ maps into the same Boolean function was calculated [1]. The results showed that for, $K \sim \mathcal{O}(1)$ when $N \gg 1$, there exists a critical average connectivity

$$K_c \approx \log_2 \log_2 \left(\frac{2N}{\ln 2}\right) + \mathcal{O}\left(\frac{1}{N \ln N}\right); \tag{3}$$

such that $\vartheta(N,K) \approx e^{\varphi N} \gg 1$ $(\varphi > 0)$ or $\vartheta(N,K) \approx 1$, depending on whether $K < K_c$ or $K > K_c$, respectively.

In genetics, NK-Kauffman networks are used as models of the genotype-phenotype map, represented by Ψ [1,2]: The genotype consists in a particular wiring, and selection of the Boolean functions f_i in (1), which give rise to the NK-Kauffman network; while the phenotype is represented by their attractors in $\Psi\left(\mathcal{L}_K^N\right) \subseteq \Lambda_N \cong \mathcal{G}_{2^N}$ [4–6]. The K connections represent the average number of epistatic interactions among the genes, and the Boolean variables S_i; the expression "1" or inhibition "0" of the i-th gene.

A well established fact in the theory of natural selection is the so-called robustness of the phenotype against mutations in the genotype [1,2,7,8]. At the level of the genotype, random mutations (by radiation in the environment) and recombination by mating, constitute the driving mechanism of the *Evolution Theory*. Experiments in laboratory controlling the amount of radiation have shown that; while the change in the phenotype vary from species to species, more than 50% of the changes have no effect at all in the phenotype [8–11]. In Ref. 1 it was shown that the signature of genetic robustness can be seen in the injective properties of Ψ, with a many-to-one map representing a robust phase. This happens if $K < K_c$, with the value of N to be substituted on (3), determined by the number of genes that living organisms have. This number varies from 6×10^3 for yeast to less than 4×10^4 in *H. sapiens* [9]. Substitution in (3) gives in both cases that $K \leq 3$ [1].

In this article it is calculated; for a general bias p, the probability $\mathcal{P}_K$ that two elements $f, f' \in \mathcal{L}_K^N$, such that f' is obtained from f by a mutation, give rise to the the same phenotype, *i.e.* $\Psi(f) = \Psi(f')$. For a mutation, it is intended a change in a Boolean function f_i, and/or its connections. The results impose restrictions in the values that K and p should have, in order that $\mathcal{P}_K \geq 1/2$, in accordance with the experiments.

The article is organized as follows: In Sec. 2, I set up a mathematical formalism that allows to write (1) in a more suitable way for calculations. In Sec. 3, the expressions of the different probabilities involved in the calculation of $\mathcal{P}_K$ are established. In Sec. 4, I introduce a new classification of Boolean functions according to its real dependence on their arguments; which I call its *degree of irreducibility*. This classification is used in Sec. 5 to calculate the invariance of Boolean functions under changes of their connections and so; calculate $\mathcal{P}_K$. In Sec. 6 the conclusions are set up. In the appendix, two errata of Ref. 1 are corrected, and it is shown that they do not alter the asymptotic results of Ref. 1. So, the biological implications there stated remain correct.

2. Mathematical Framework

Now we introduce a mathematical formalism with the scope of write (1) in a more suitable notation for counting. All additions between elements of $\mathbb{Z}_2$ and its cartesian products are modulo 2.

Let $\mathcal{M}_N = \{1, 2, \ldots, N\}$ denote the set of the first N natural numbers. A K-connection set C_K, is any subset of $\mathcal{M}_N$ with cardinality K. Since there are $\binom{N}{K}$ possible K-connection sets; we count them in some unspecified order, and denote them by

$$C_K^{(\alpha)} = \{i_1, i_2, \ldots, i_K\} \subseteq \mathcal{M}_N, \text{ with } \alpha = 1, \ldots, \binom{N}{K}, \tag{4}$$

where, without lost of generality; $i_1 < i_2 < \cdots < i_K$, with $1 \leq i_m \leq N$ $(1 \leq m \leq K)$. We also denote as

$$\Gamma_K^N = \left\{ C_K^{(\alpha)} \right\}_{\alpha=1}^{\binom{N}{K}} \tag{5}$$

the set of all K-connections. To each K-connection set $C_K^{(\alpha)}$ it is possible to associate a K-connection map

$$C_K^{*(\alpha)} : \mathbb{Z}_2^N \longrightarrow \mathbb{Z}_2^K,$$

defined by

$$C_K^{*(\alpha)}(\mathbf{S}) = C_K^{*(\alpha)}(S_1, \ldots, S_N) = (S_{i_1}, \ldots, S_{i_K}) \quad \forall\, \mathbf{S} \in \mathbb{Z}_2^N.$$

Any map

$$b_K : \mathbb{Z}_2^K \to \mathbb{Z}_2, \tag{6}$$

defines a K-Boolean function. Since $\#\mathbb{Z}_2^K = 2^K$; b_K is completely determined by its K-truth table T_K, given by

$$T_K = [A_K \ \mathbf{b}_K],$$

where A_K is a $2^K \times K$ binary matrix, and $\mathbf{b}_K$ is a 2^K dimensional column-vector, such that:

The s-th row $(1 \leq s \leq 2^K)$ of A_K encodes the binary decomposition of s, and represents each one of the possible 2^K arguments of b_K in (6). So, A_K satisfies [1]

$$s = 1 + \sum_{i=1}^{K} A_K(s, i)\, 2^{i-1}.$$

And

$$\mathbf{b}_K = [\sigma_1, \sigma_2, \ldots, \sigma_{2^K}]\,, \tag{7}$$

where $\sigma_s \in \mathbb{Z}_2$ $(1 \leq s \leq 2^K)$, represents the images of (6).

There are as much as 2^{2^K} K-truth tables T_K corresponding to the total possible vectors (7). K-Boolean functions can be classified according to Wolfram's notation by their decimal number μ given by [1,12]

$$\mu = 1 + \sum_{s=1}^{2^K} 2^{s-1}\sigma_s. \tag{8}$$

Let us add a superscript (μ) to a K-Boolean function, or to its truth table, whenever we want to identify them. So, $b_K^{(\mu)}$ and $T_K^{(\mu)}$ refer to the μ-th K-Boolean function and its truth table respectively. Within this notation, the set of all K-Boolean functions Ξ_K is expressed as

$$\Xi_K = \left\{ b_K^{(\mu)} \right\}_{\mu=1}^{2^{2^K}}.$$

Of particular importance are the *tautology* $b_K^{(\tau)} \equiv b_K^{(2^{2^K})}$ and *contradiction* $b_K^{(\kappa)} \equiv b_K^{(1)}$ K-Boolean functions; with images:

$$\mathbf{b}_K^{(\tau)} = [1, 1, \ldots, 1] \tag{9a}$$

and

$$\mathbf{b}_K^{(\kappa)} = [0, 0, \ldots, 0]\,. \tag{9b}$$

Table 1 gives an example of the A_2 matrix (representing the four possible entries of S_1 and S_2) with the sixteen possible 2-Boolean functions listed according to their decimal number (8).

S_1	S_2	$\mu\ \mapsto$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
0	0	$\sigma_1 \mapsto$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1	0	$\sigma_2 \mapsto$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	1	$\sigma_3 \mapsto$	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	1	$\sigma_4 \mapsto$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

Table 1. The A_2 matrix, with the sixteen 2-Boolean functions.

Within the preceding notation, the dynamical rule (1), now may be rewritten as

$$S_i\,(t+1) = b_K^{(\mu_i)} \circ C_K^{*(\alpha_i)}\,(\mathbf{S}\,(t))\,, \quad i = 1, \ldots, N; \tag{10}$$

where, some of the indexes α_i and μ_i may be equal for different values of i, and $\mathbf{S}\,(t) \in \mathbb{Z}_2^N$.

3. The Invariance of NK-Kauffman Networks

Now we are interested in calculate the probability $\mathcal{P}_K$, that (10) remains invariant under a change of a connection $C_K^{(\alpha)}$ and/or a K-Boolean function $b_K^{(\mu)}$. So, we must study the number of ways in which this could happen; *i.e.* what conditions should prevail in order that for some i,

$$b_K^{(\mu_i)} \circ C_K^{*(\alpha_i)}\,(\mathbf{S}) + b_K^{(\nu_i)} \circ C_K^{*(\beta_i)}\,(\mathbf{S}) \;=\; 0 \quad \forall\ \mathbf{S} \in \mathbb{Z}_2^N, \tag{11a}$$

for $\alpha_i \neq \beta_i$ and/or $\mu_i \neq \nu_i$. Let us use a shorthand notation and skip to write the indexes α_i, and μ_i. Then (11a) may be written as

$$\tilde{b}_K \circ \tilde{C}_K^*\,(\mathbf{S}) + b_K \circ C_K^*\,(\mathbf{S}) = 0 \quad \forall\ \mathbf{S} \in \mathbb{Z}_2^N,$$

where, $\tilde{b}_K = b_K + \Delta b_K$ and $\tilde{C}_K^* = C_K^* + \Delta C_K^*$; with $\Delta b_K \in \Xi_K$, and ΔC_K^* a K-connection map. Explicit substitution gives

$$b_K \circ \Delta C_K^*\,(\mathbf{S}) + \Delta b_K \circ \tilde{C}_K^*\,(\mathbf{S}) = 0 \quad \forall\ \mathbf{S} \in \mathbb{Z}_2^N. \tag{11b}$$

Equation (11b) could be satisfied in three different ways:

i) Event $\mathcal{A}$: A change in a K-connection C_K without a change in a K-Boolean function b_K. This implies $\Delta b_K = 0\ \forall\,\mathbf{S} \in \mathbb{Z}_2^K$, and from (9b) $\Rightarrow b_K \circ \Delta C_K^*\,(\mathbf{S}) = b_N^{(\kappa)}$.

ii) Event $\mathcal{B}$: A change in a K-Boolean function b_K without a change in a K-connection C_K. This implies $\Delta C_K^* = 0 \; \forall \, \mathbf{S} \in \mathbb{Z}_2^N$. From (9b) there follows $\Delta b_K \circ \tilde{C}_K^* (\mathbf{S}) = 0 \; \forall \, \mathbf{S} \in \mathbb{Z}_2^N \Rightarrow \Delta b_K = b_K^{(\kappa)}$. So, the K-Boolean function must remain unchanged.

iii) Event $\mathcal{C}$: A change in a K-Boolean function b_K and a change in a K-connection C_K. In this case, both $\Delta C_K^* \neq 0$ and $\Delta b_K \neq 0$; and also (11b) must hold with independency of $\mathcal{A}$, and $\mathcal{B}$ events. So, from (9a), it must happen that $b_K \circ \Delta C_K^* = b_N^{(\tau)}$, and $\Delta b_K \circ \tilde{C}_K^* = b_N^{(\tau)}$.

Since the events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, are independent, the probability $\mathcal{P}_K$ that (11) are satisfied, is given by the combined probabilities $P(\mathcal{A})$, $P(\mathcal{B})$, and $P(\mathcal{C})$ that $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ happen. So,

$$
\begin{aligned}
\mathcal{P}_K &= P(\mathcal{A}) + P(\mathcal{B}) + P(\mathcal{C}) - P(\mathcal{A})\; P(\mathcal{B}) - P(\mathcal{A})\; P(\mathcal{C}) - P(\mathcal{B})\; P(\mathcal{C}) \\
&+ P(\mathcal{A})\; P(\mathcal{B})\; P(\mathcal{C})\,. \qquad (12)
\end{aligned}
$$

For a general bias p $(0 < p < 1)$ that $\sigma_s = 1$, for $1 \leq s \leq 2^K$ in (7), the probability $\Pi(b_K)$ to extract the K-Boolean function b_K is given by

$$
\Pi(b_K) = p^{\omega} (1-p)^{2^K - \omega} \quad , \qquad (13a)
$$

where

$$
\omega = \omega(b_K) = \sum_{s=1}^{2^K} \sigma_s, \qquad (13b)
$$

is the *weight* of b_K.

The following considerations are in order:

i) $P(\mathcal{A})$ is the probability that the *projected function*

$$
b_K^{*(\alpha)} \equiv b_K \circ C_K^{*(\alpha)} : \mathbb{Z}_2^N \to \mathbb{Z}_2 \qquad (14)
$$

remains invariant under a change of the K-connection. To get read of this we must first introduce the concept of irreducibility of Boolean functions, which is going to be done in the next section.

ii) $P(\mathcal{B})$ is the average probability that b_K remains invariant by a mutation, given that b_K has occurred. Then

$$
P(\mathcal{B}) = \sum_{b_K \in \Xi_K} \Pi^2(b_K)\,.
$$

Since there are $\binom{2^K}{\omega}$ K-Boolean functions with weight ω, from (13) we obtain

$$P(\mathcal{B}) = \sum_{\omega=0}^{2^K} \binom{2^K}{\omega} p^{2\omega} \ (1-p)^{2^{K+1}-2\omega} = [1-2p(1-p)]^{2^K}.$$

iii) $P(\mathcal{C})$ is the probability of extracting twice the *tautology* N-Boolean function. From (9a) and (13b) $\omega\left(b_N^{(\tau)}\right) = 2^N$, so from (13a)

$$P(\mathcal{C}) = \Pi^2\left(b_N^{(\tau)}\right) = p^{2^{N+1}} \ll 1.$$

So, we obtain the asymptotic expression for (12)

$$\mathcal{P}_K \approx P(\mathcal{A}) + [1-2p(1-p)]^{2^K}[1-P(\mathcal{A})] + \mathcal{O}\left(p^{2^{N+1}}\right), \qquad (15)$$

for $N \gg 1$.

4. The Irreducibility of the Boolean Functions

Not all the K-Boolean functions depend completely on their K arguments. For instance, let us consider the 2-Boolean functions of table 1: Rules **1** and **16** (*contradiction* and *tautology*, respectively) do not depend on either S_1 or S_2; while rules **6** and **11** (*negation* and *identity*, respectively) only depend on S_1. Due to this fact, let us make the following definitions:

Definition 1

A K-Boolean function b_K is *reducible* on the m-th argument S_m ($1 \leq m \leq K$), if

$$b_K(S_1, \ldots, S_m, \ldots, S_K) = b_K(S_1, \ldots, S_m + 1, \ldots, S_K) \quad \forall\, \mathbf{S} \in \mathbb{Z}_2^K.$$

Otherwise, the K-Boolean function b_K is *irreducible* on the m-th argument S_m.

Definition 2

A K-Boolean function b_K is *irreducible of degree* λ $(0 \leq \lambda \leq K)$; if it is irreducible on λ arguments and reducible on the remaining $K-\lambda$ arguments. If $\lambda = K$, the K-Boolean function is *irreducible.*

Let us denote by $\mathcal{I}_K(\lambda)$ the set of irreducible K-Boolean functions of degree λ. From definitions 1 & 2, Ξ_K may be decomposed uniquely in terms of $\mathcal{I}_K(\lambda)$ by

$$\Xi_K = \bigcup_{\lambda=0}^{K} \mathcal{I}_K(\lambda)\,, \tag{16a}$$

with

$$\mathcal{I}_K(\lambda) \cap \mathcal{I}_K(\lambda') = \emptyset \quad \text{for} \quad \lambda \neq \lambda'. \tag{16b}$$

The cardinalities $\beta_K(\lambda) \equiv \#\mathcal{I}_K(\lambda)$ may be calculated recursively, noting that $\beta_K(\lambda)$, must be equal to the number of ways to form λ irreducible arguments from K arguments. This amounts to $\binom{K}{\lambda}$ times the number of irreducible λ-Boolean functions $\beta_\lambda(\lambda)$; thus

$$\beta_K(\lambda) = \binom{K}{\lambda} \beta_\lambda(\lambda)\,. \tag{17}$$

Setting $K = \lambda$ in (16a) and calculating the cardinalities, follows that

$$2^{2^\lambda} = \sum_{\nu=0}^{\lambda-1} \beta_\lambda(\nu) + \beta_\lambda(\lambda)\,.$$

Substituting back into (17) the following recursion formulas for the number of irreducible K-Boolean functions of degree λ are obtained

$$\beta_K(\lambda) = \binom{K}{\lambda} \left[2^{2^\lambda} - \sum_{\nu=0}^{\lambda-1} \beta_\lambda(\nu)\right], \tag{18a}$$

and

$$\beta_K(0) = 2. \tag{18b}$$

Note from (9), that $b_K^{(\tau)}$ and $b_K^{(\kappa)}$ are irreducible of degree zero. So from (18b),

$$\mathcal{I}_K(0) = \left\{b_K^{(\tau)},\, b_K^{(\kappa)}\right\}. \tag{19}$$

Some first values for $\beta_K(\lambda)$ are:

$$\beta_K(1) = 2K,$$
$$\beta_K(2) = 5K(K-1),$$
$$\beta_K(3) = \frac{109}{3}K(K-1)(K-2),$$
$$\beta_K(4) = \frac{32,297}{12}K(K-1)(K-2)(K-3),$$

etc.

5. The Probability $P(\mathcal{A})$

Let us now calculate $P(\mathcal{A})$ to obtain $\mathcal{P}_K$ from (15). The probability $P(\mathcal{A})$, that $b_K^{*(\alpha)}$, defined by (14), remains invariant against a change in $C_K^{(\alpha)}$, depends in the degree of irreducibility of b_K; *i.e.* on which of its K arguments it really depends. To calculate it, let us first calculate the probability $P\left[\Delta\; b_K^{*(\alpha)} = 0 | b_K \in \mathcal{I}_K(\lambda)\right]$ that, $b_K^{*(\alpha)}$ remains invariant due to a change in the K-connection $C_K^{(\alpha)}$; given that b_K is irreducible of degree λ.

Let $b_K \in \mathcal{I}_K(\lambda)$ be irreducible in the arguments with indexes

$$m_1, m_2, \ldots, m_\lambda\ , \text{ where } \ m_1 < m_2 < \cdots < m_\lambda$$

such that $1 \leq m_l \leq K$ $(1 \leq l \leq \lambda)$. Let us also rewrite (4) more explicitly putting the superscript (α) into its elements; then

$$C_K^{(\alpha)} = \left\{ i_1^{(\alpha)}, i_2^{(\alpha)}, \ldots, i_K^{(\alpha)} \right\} \subseteq \mathcal{M}_N.$$

Now, associated to $b_K^{*(\alpha)}$, we can define its λ-*irreducible connection* by

$$\mathcal{J}_\lambda\left(b_K^{*(\alpha)}\right) \equiv \left\{ i_{m_l}^{(\alpha)} \right\}_{l=1}^{\lambda} \subseteq C_K^{(\alpha)}.$$

Within this notation the set $\Theta_K^N\left(b_K^{*(\alpha)}\right)$, of the K-connections $C_K^{(\beta)}$ that leave $b_K^{*(\alpha)}$ invariant, is given by

$$\Theta_K^N\left(b_K^{*(\alpha)}\right) = \left\{ C_K^{(\beta)} \in \Gamma_K^N \mid i_{m_l}^{(\beta)} = i_{m_l}^{(\alpha)}\ \forall\ l = 1, 2, \ldots, \lambda \right\}. \qquad (20)$$

Then

$$P\left[\Delta\, b_K^{*(\alpha)} = 0 | b_K \in \mathcal{I}_K(\lambda)\right] = \frac{\#\Theta_K^N\left(b_K^{*(\alpha)}\right)}{\#\Gamma_K^N}\ . \tag{21}$$

From (5), $\#\Gamma_K^N = \binom{N}{K}$. To calculate $\#\Theta_K^N\left(b_K^{*(\alpha)}\right)$, let us note that the K-connections $C_K^{(\beta)} \in \Theta_K^N\left(b_K^{*(\alpha)}\right)$ have λ elements fixed, the elements of $\mathcal{J}_\lambda\left(b_K^{*(\alpha)}\right)$, and $K-\lambda$ elements free, which are the elements of $\mathcal{M}_N \setminus \mathcal{J}_\lambda\left(b_K^{*(\alpha)}\right)$. Thus, $\#\Theta_K^N\left(b_K^{*(\alpha)}\right)$ equals the number of subsets of $\mathcal{M}_N \setminus \mathcal{J}_\lambda\left(b_K^{*(\alpha)}\right)$ that can be constructed with $K-\lambda$ elements. Since

$$\#\left[\mathcal{M}_N \setminus \mathcal{J}_\lambda\left(b_K^{*(\alpha)}\right)\right] = N - \lambda,$$

we obtain

$$\#\Theta_K^N\left(b_K^{*(\alpha)}\right) = \binom{N-\lambda}{K-\lambda}. \tag{22}$$

That only depends in the degree of irreducibility λ of b_K and not in the connection index (α). Substituting (22) into (21) we obtain

$$P\left[\Delta\, b_K^{*(\alpha)} = 0 | b_K \in \mathcal{I}_K(\lambda)\right] = \frac{K!\,(N-\lambda)!}{N!\,(K-\lambda)!}\ . \tag{23a}$$

Due to (16), $P(\mathcal{A})$ is given by:

$$P(\mathcal{A}) = \sum_{\lambda=0}^{K} P\left[\Delta\, b_K^{*(\alpha)} = 0 | b_K \in \mathcal{I}_K(\lambda)\right]\ P\left[b_K \in \mathcal{I}_K(\lambda)\right], \tag{23b}$$

where $P\left[b_K \in \mathcal{I}_K(\lambda)\right]$ is the probability that b_K be irreducible of degree λ. The value of $P\left[b_K \in \mathcal{I}_K(\lambda)\right]$ depends on $\beta_K(\lambda)$ [calculated from (18)], as well as on the particular way in which the K-Boolean functions b_K are extracted.

When $K \sim \mathcal{O}(1)$ for $N \gg 1$, equations (23) behave asymptotically like

$$P(\mathcal{A}) \approx P\left[b_K \in \mathcal{I}_K(0)\right] + \mathcal{O}\left(\frac{1}{N}\right).$$

So from (19), the leading term of $P(\mathcal{A})$ comes from the probability to extract the *tautology* (9a) and *contradiction* (9b) K-Boolean functions. We obtain from (13)

$$P(\mathcal{A}) \approx p^{2^K} + (1-p)^{2^K} + \mathcal{O}\left(\frac{1}{N}\right).$$

From (15) the probability that (1) [or equivalently (10)] remains invariant by a change on a K-Boolean function and/or its connection; is given by

$$\begin{aligned} \mathcal{P}_K &\approx p^{2^K} + (1-p)^{2^K} \\ &+ [1-2p(1-p)]^{2^K} \left\{1 - \left[p^{2^K} + (1-p)^{2^K}\right]\right\} + \mathcal{O}\left(\frac{1}{N}\right). \end{aligned} \tag{24}$$

6. Conclusion

A classification of K-Boolean functions in terms of its irreducible degree of connectivity λ was introduced. This allowed us to uniquely decompose them through (16), and calculate the asymptotic formula (24) for $\mathcal{P}_K$; that an NK-Kauffman network (1) remains invariant against a change in a K-Boolean function and/or its K-connection. Figure 1 shows the graphs for $\mathcal{P}_K$ *vs* p; for different values of the average connectivity K. The graphs attain a minimum and are symmetric at $p = 1/2$ (the case of a uniform distribution). For p fixed, $\mathcal{P}_K < \mathcal{P}_{K'}$ for $K > K'$.

These results are specially important when NK-Kauffman network are used to model the genotype-phenotype map (2) [1,2]. Experiments to study the robustness of the genetic material have been done by means of induced mutations [9–11]. The results varied among the different organisms studied, but it is estimated that in more than 50% of the cases the phenotype appears not to be damaged. In NK-Kauffman networks this phenomena is manifest when $\mathcal{P}_K > 1/2$. Figure 1 shows that is possible to be in agreement with the experimental data without a bias ($p = 1/2$), provided $K \leq 1.25$ for the average connectivity. For the case $K = 2$ this happens only for values of p outside the interval $[0.21, 0.78]$. There is no surprise that biassed values of p increment the value of $\mathcal{P}_K$ since they tend to increase the amount of *tautology* and *contradiction* functions (9) through (13).

Acknowledgments

This work is supported in part by **CONACyT** project number **059869** and **PAPIIT** project number **IN101309-3**. The author wishes to thank: Martha Takane for fruitful mathematical discussions, Thalía Figueras for careful reading of the manuscript, Mamed Atakishiyev for computational advice and Pilar López Rico for accurate services on informatics.

Appendix: Errata in Ref. 1

All quotations to equations in Ref. 1 are preceded by an "R", those introduced here by an "A", while all the others refer to equations of the present article.

In Ref. 1 it was wrongly stated that the only Boolean functions that contribute to the number of redundances r in (R16) are: the *tautology*, the *contradiction*, the *identity* and the *negation*. In fact there are contributions from many more functions, their number growing with K for $K < N$ (in the case $K = N$ of the *random map model* $r = 0$; as explained further); according to their classification in terms of its degree of irreducibility defined in Sec. 4 of this article. Furthermore; the contribution to r of the *identity* and *negation* functions were calculated as $2N\left[\binom{N-1}{K-1} - 1\right]$, while the correct value is

$$2K\left[\binom{N-1}{K-1} - 1\right]. \tag{A1}$$

Nevertheless these inconveniences:

- In the asymptotic expansion of (R18) for $N \gg 1$, the contribution $\mathcal{O}(1)$ is originated from the *tautology* and *contradiction* functions.
- While the wrong reported contribution $2N\left[\binom{N-1}{K-1} - 1\right]$, of the *identity* and *contradiction* functions, turns out to be $\mathcal{O}(1)$, it just adds an extra term $\ln(K_c + 1)$ in (R22) that does not contribute to the $\mathcal{O}(1)$ term of its solution (R23). However it gives a wrong, and slower, decaying error $\mathcal{O}(\ln\ln\ln N/\ln N)$.
- The rest of the Boolean functions, with $\lambda \geq 2$, give an $\mathcal{O}(1/N^2)$ contribution to (R18).

This implies that all the asymptotic results and their genetical consequences remain correct; while the decaying error term in (R23) becomes $\mathcal{O}(1/N\ln N)$ since the correct value (A1) gives a contribution $\mathcal{O}(1/N)$ to (R18).

The correct results are obtained as follows:

From (18) and (20), the number of redundances that the elements of $\mathcal{I}_K(\lambda)$ furnish is given by $\beta_K(\lambda)\left[\#\Theta_K^N(\lambda) - 1\right]$. From (22), the correct value of r is:

$$r = \sum_{\lambda=0}^{K} \beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]. \tag{A2}$$

Note that:

- The contribution of $\lambda = 0$, is the one that corresponds to the *tautology* and *contradiction* K-Boolean functions.
- The contribution of $\lambda = 1$, is the one given by (A1), with $\beta_K(1) = 2K$ obtained from (18).
- The contribution of $\lambda = K$ is zero. So, irreducible K-Boolean functions give raise to injective maps.
- In the special case of the *random map model* [3,5,13]: $r = 0$ as it should be, due to the fact that, for such a case $\Psi : \mathcal{L}_N^N \to \Lambda_N$ defined by (R4) [respectively by (2) in this article], becomes a bijection so

  $$\mathcal{L}_N^N \cong \Xi_N \cong \mathcal{G}_{2^N},$$

  where $\mathcal{G}_{2^N}$ is the set of functional graphs from 2^N points to themselves [1].

With this background, the correct equations (R17), (R18), (R19), (R22), (R23), and (R25); are given as follows:

From (R16) and (A2) we obtain

$$\#\Psi\left(\mathcal{L}_K^N\right) = \left\{2^{2^K}\binom{N}{K} - \sum_{\lambda=0}^{K}\beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]\right\}^N. \tag{R17}$$

Now

$$\vartheta^{-1}(N,K) = \{1 - \varphi(N,K)\}^N, \tag{R18}$$

with φ depending also on N; and given by

$$\varphi(N,K) = \frac{\sum_{\lambda=0}^{K}\beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]}{2^{2^K}\binom{N}{K}}. \tag{R19}$$

From (18), $\varphi(N,K)$ admits for $N \gg 1$ the asymptotic expansion

$$\varphi(N,K) \approx \frac{1}{2^{2^K-1}}\left[1 + \mathcal{O}\left(\frac{1}{N}\right)\right];$$

that gives for the equation $\vartheta^{-1}(N, K_c) = 1/2$, of the critical connectivity,

$$2^{2^{K_c}} \approx \frac{2N}{\ln 2}\left[1 + \mathcal{O}\left(\frac{1}{N}\right)\right]. \tag{R22}$$

The solution of (R22) is now

$$K_c \approx \log_2 \log_2 \left(\frac{2N}{\ln 2} \right) + \mathcal{O}\left(\frac{1}{N \ln N} \right). \tag{R23}$$

And (R25) is now given by

$$\Delta K_c \approx \frac{2}{(\ln 2)^3 \ \log_2 (2N/\ln 2)} \sim \mathcal{O}\left(\frac{1}{\ln N} \right). \tag{R25}$$

This shows that the asymptotic formulas, and conclusions of Ref. 1 are correct.

References

[1] Romero, D., and Zertuche, F., *Number of Different Binary Functions Generated by NK-Kauffman Networks and the Emergence of Genetic Robustness*. J. Math. Phys. **48** (2007) 083506.

[2] Wagner A., *Does Evolutionary Plasticity Evolve?* Evolution **50** (1996) 1008-1023.

[3] Romero, D., and Zertuche, F., *The Asymptotic Number of Attractors in the Random Map Model*. J. Phys. A: Math. Gen. **36** (2003) 3691; *Grasping the Connectivity of Random Functional Graphs*. Stud. Sci. Math. Hung. **42** (2005) 1.

[4] Kauffman, S.A., *Metabolic Stability and Epigenesis in Randomly Connected Nets*. J. Theoret. Biol. **22** (1969) 437.

[5] Kauffman, S.A., *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press (1993).

[6] Kauffman, S.A. *The Large-Scale Structure and Dynamics of Gene Control Circuits: An Ensemble Approach*. J. Theoret. Biol. **44** (1974) 167; *Developmental Logic and its Evolution*. BioEssays **6** (1986) 82; *A Framework to Think about Regulatory Systems*. In: Integrating Scientific Disciplines. (Ed. W. Bechte) (1986) Martinus Nijhoff, Dordrecht.

[7] Wagner, A., *Robustness and Evolvability in Living Systems*. Princeton University Press (2005).

[8] de Visser J.A.G.M., *et.al.*, *Perspective: Evolution and Detection of Genetic Robustness*. Evolution **57** (2003) 1959-1972.

[9] Lewin, B., *GENES IX*. Jones and Bartlett Pub. (2008).

[10] Goebel, M.G. and Petes, T.D., *Most of the Yeast Genomic Sequences are not Essential for Cell Growth and Division*. Cell **46** (1986) 983-992; Hutchison, C.A. *et. al.*, *Global Transposon Mutagenesis and a Minimal Mycoplasma Genome*. Science **286** (1999) 2165-2169; Giaver, G. *et. al.*, *Functional profiling of the* S. cerevisiae *genome*. Nature **418** (2002) 387-391.

[11] Thatcher, J.W., Shaw, J.M., and Dickinson, W.J. *Marginal Fitness Contributions of Nonessential Genes in Yeast*. Proc. Natl. Acad. Sci. USA **95** (1998) 253-257.

[12] Weisbuch, G., *Complex Systems Dynamics.* Addison Wesley, Redwood City, CA (1991); Wolfram, S., *Universality and Complexity in Cellular Automata.* Physica D **10** (1984) 1.

[13] Derrida, B., and Flyvbjerg, H., *The Random Map Model: a Disordered Model with Deterministic Dynamics.* J. Physique **48** (1987) 971.

Figure caption:

Figure 1. (Color online), Graphs for $\mathcal{P}_K$ *vs.* p for different values of the average connectivity K. $K = 1$ in red, $K = 1.25$ in green and $K = 2$ in blue. The important $\mathcal{P}_K = 1/2$ value, is in magenta.